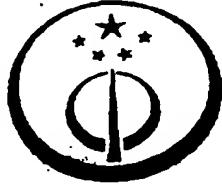


Abstract

Method of Searching Chinese Word Indexed with Chinese Character Stroke Shapes and Processing Method therefor

A method of searching Chinese word indexed with stroke shapes is introduced, in which a Chinese word is searched by the type(s) of its first or first to second start strokes, and the searching domain is reduced by indexing Chinese words with the strokes categorized into horizontal, vertical, oblique, dot-like, crooked ones and their number codes. Such a method is implemented in a computer having the compatibility of Chinese and English, as an application in an operating system. For example, the application is programmed in C language and the operating system is UNIX operating system, wherein the memory addresses of Chinese characters are sorted according to the amount of their strokes.



〔12〕发明专利申请公开说明书

[11] CN 85 1 05023 A

[43]公开日 1987年1月7日

(21)申请号 85 1 05023

(22)申请日 85.7.2

·(71)申请人 张 楠

地址 美国纽约11368.科罗纳.阿普特8V.第
99街52-09号

(72)发明人 张 楠

(74)专利代理机构 中国国际贸易促进委员会专利

代理部

代理人 董江雄

(54)发明名称 汉字笔形检字编码方法及其处理方法

(57) 摘要

采用笔形检字编码法,按汉字起笔的第一或第一、第二笔的笔划种类检字,并用横、直、斜、点、折所归纳的五种笔划及其数字代码把汉字重新分类以缩小检字范围。笔形检字编码法以操作系统的实用程序与中、英文兼用计算机相联系,采用C语言和UNIX操作系统,其中的汉字内存码按照汉字笔划的多少来排列。

1. 一种汉字编码方法。其特征为：采用笔形检字编码法，按汉字起笔的第一或第一、第二笔的笔划种类检字。把汉字重新分类以缩小检字范围。

2. 根据权利要求1的笔形检字编码法，可把缩小范围后的字组显示在屏幕上，以便在少数的字组中指定要找的单字。其特征为将起笔笔划归纳为横、直、斜、点、折五种笔划，并分别用一个数码代表，且对应输入键盘的五个键。

3. 根据权利要求1或2的笔形检字编码法及其笔划归纳。其特征为所说的五种笔划与数码的对应关系为：

横 = 1 直 = 2 斜 = 3 点 = 4 折 = 5

这里的五个数码分别对应输入键盘的 1, 2, 3, 4, 5 五个键。

4. 一种汉字分类排列的编码。其特征为是按汉字笔划多少进行排列。可用现有的“分类程序”把中文资料按笔划多少进行排列。

汉字笔形检字编码方法及其处理方法

本发明所涉及的是汉字笔形检字编码方法及其处理方法。更具体讲，是涉及了一种简单易学的汉字编码方法及其计算机输入汉字的方法。并且是以操作系统的实用程序来实现的。

目前，国内的汉字编码方法及其处理方法有许多种。一般说来，这些汉字编码方法及其处理方法往往是：

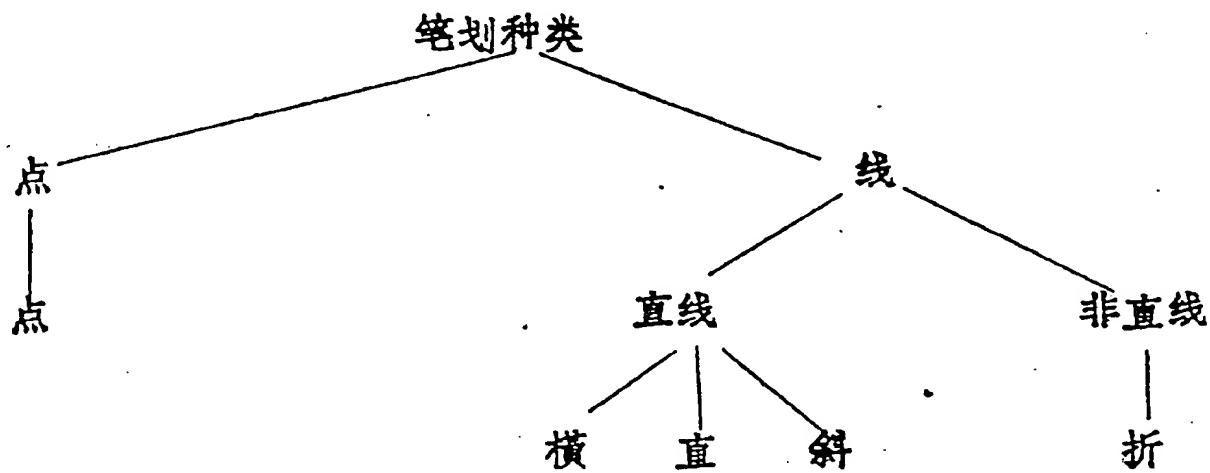
1. 采用繁杂的检字法输入汉字，使用时需要特殊的训练；
2. 所用的输入键盘，同一键钮代表多种符号，难于记忆；而且所用的键钮较多，不便操作；
3. 采用汇编语言的计算机程序，只能用于特定的计算机；
4. 把中文字库做在终端机的硬件上，缺乏弹性，也不经济。

本发明的目的是采用汉字笔形检字编码方法，实现编码规则简单易用，所用字键很少，没有经过训练的非专职人员也能够很快掌握，另外在计算机硬件的配合上具有较大弹性，不局限于某一特定的计算机。

中文计算机的输出。主要的是打印汉字。所有的中文计算机都用点阵表现汉字。并无特殊之处，所不同的只是输入汉字。所谓输入汉字。一般的说。只是在预存的字库里指定一个单字。换句话说。只是选择一种检字法。本发明所采用的“笔形检字编码法”是改良的部首检字法。依照起笔的第一或第一、第二笔划种类检字。不再数笔划。

明以汉字的笔划分类为基础。所以首先要对笔划给予定义。即使笔划名称相同。但是定义不同。用法也不同。本发明的笔划种类是根据几何学的观点。把汉字的笔划归纳为横、直、斜、点、折五种。因为汉字也是

一种平面的几何图形。平面的图形只有点和线两种。线又分直线和非直线两种。直线再分横、直、斜三种。共为五种。用图解法表示更可以看出它的完整性和概括性。



这些笔划是以手写字体为准。如果以印刷字体为准也未尝不可。这只是约定问题。与方法无关。同时。所谓直线也不是严格的直线。尤其是撇捺都带一点弧形。但是在书法的观念上总是当它直线。以下是各种笔划的定义：

横 包括任何水平式的笔划。不论长短。

直 包括任何垂直式的笔划。不论带勾或不带勾。也不论长短。

斜 包括任何非横非直的直线式的笔划。不论撇捺。也不论长短。

点 包括任何形式的点。例如“心”字中的各种点。

折 包括任何非点非直线的笔划。有明显的角度和转折。包括带勾的笔划。不论大小。

以上的笔划都有一个代表号码。现约定其号码如下：

横 = 1 直 = 2 斜 = 3 点 = 4 折 = 5

检字法的目的都是逐步缩小检字的范围。最后只剩下一个字。就是要找的。“笔形检字编码法”可以用在部首检字法。也可以用在拼音检字法或其他的检字法。同样可以用来缩小检字的范围。最终指定了某个单字。以下就结合附图说明怎样把“笔形检字编码法”应用在中文计算机的汉字输入上。

图1为“笔形检字编码法”的部首编码表。

用“笔形检字编码法”“开中之是一并不正确的插入汉字的方法”与传统的部首检字法加以改良也可以达到同样的目的。部首检字法有两个缺点：一是有些字难定部首。二是要数笔划。前者可把同一个字列在几个可能的部首。后者可以用“笔形检字编码法”解决。

采用“笔形检字编码法”不必数笔划。只看起笔的笔划种类。查部首看笔划。查单字也看笔划。查部首利用部首编码表。见附图一。它把所有的部首按起笔头两笔的笔划种类分为二十三类。确定起笔后。很容易就找到部首号码。再将部首分类号码加各个部首下所注的号码，在计算机处理中文的时候。屏幕上会出现“#”记号。意思要打进部首号码。指定部首。

指定部首之后。可能有以下三种情形：

1. 字数很少的部首。例如“建”字在“乚”部。起笔的笔画是折斜。在部首编码表的折斜类。很快就找到“乚”部。它的部首编码是5 3 4。打进这个号码就指定了“乚”部。“乚”部的字很少。所以屏幕上就会自动显示所有“乚”的字如“亼”：

廷	延	建
1	2	3

屏幕上每字下都有一个号码。建字的号码是3。打进3就指定建字。

2. 字数不多也不少的部首。例如“市”字在中部。起笔是直折。部首号码是233。打进这个号码。屏幕上会出现一个问号“？”。意思是说要打进一个起笔的笔画号码（部首的笔画不计）。市字的起笔是点。号码是4。打进4。中部点起笔的字就显在屏幕上如下：

市 帝 帛 帐
1 2 3 4

打进1就指市字。

3. 字数很多的部首。例如“艸”字在艸部。起笔是横直。部首号码是122。打进号码。屏幕上会显示两个问号“？？”。意要打进“艸”字的二个起笔号码。“艸”字的起笔是直折。打进25。屏幕就显示如下：

口字 萸 萸 萸 萸 萔
1 2 3 4 5 6

芮 萸 萸 萔 萔 萔 萔
7 8 9 10 11 12 13 14 15

日曰 朞 朞 朞 朞 朞
16 17 18 19 20 21 22

四目 萝 萝 萝 萝 萝
23 24 25 26 27 28 29

艹艸 莫 莫 莫 莫 莫 莫
30 31 32 33 34 35 36 37

从以上的实例中，可以知道汉字的笔画很多，如果按笔画的多少来分形再分类，才便于查看。这些小类的名称和赋予的意义都非常明显。照这些类名来看，幕字一定在“莫字”类，不必再看别类的字。很快就从“莫字”类找到幕字，打进3,2就指定幕字。

附带声明的是部首编码表是作为参考，专供初学者使用，稍微熟练之后，至少可以记住常用的部首号码，不必经常查看部首编码表。

从以上的实例中，可以得到“笔形检字编码”的定义。所谓“笔”，是说指定部首和指定单字都以起笔的笔画为准。所谓“形”，是说有时按照部分相同的字分类。本发明所称的“笔形”是一个特别名词，“笔形检字法”也是一个特别名词，与一般的用法不同。

○语言和 UNIX 操作系统是分不开的，它们的优点已有定论，不必多说。本发明选用 C 语言和 UNIX 操作系统，是因为可以不受计算机硬件的限制，对日新月异的计算机硬件的发展，更具有高度的适应性。

本发明通过在 UNIX 操作系统中增设一些程序，就可使“笔形检字编码法”输入汉字得到实际应用。它是操作系统下的处理中文的实用程序。增设这些程序，就可以把支持 C 语言和 UNIX 操作系统的计算机改变为中英文兼用的计算机。作为中文计算机，只要在 UNIX 操作系统下增加四个处理中文的函数：

1、 `uscanf` 这是 `scanf` 从键盘上输入汉字的函数，相当 UNIX 操作系统原有的 `scanf`，也即从键盘输入英文字母的函数。

2、 `cprintf` 这是把汉字显示在屏幕上的函数，相当原有的 `printf`，也即把字母显示在屏幕的函数。

3、 `cfprintf` 这是把汉字储存到指定档案的函数，相当原有的

`fprintf`，也即把字母储存到指定档案的函数。

4、`C I P` 这是打印汉字的函数，相当原有的 `lpr`，也即打印字母的函数。

采用 24×24 的点子方阵，每个单字要 72 个字节。即使采用 16×16 的点子方阵，每个单字也要 32 个字节。所以内存大都采用四位数的代字号码。中华人民共和国的国家标准信息交换用的汉字编码就是一种标准内存。本发明依照汉字笔画多少排列的四位数编码。这种编码的好处是可以用现有的“分类程序”把中文资料按笔划多少排列，就像英文可照字母顺序排列。这种按笔划多少排列的内存码，定名为“张氏汉字编码”简称“张码”。张码和国标码有交叉参考关系，与国标码并无抵触之处。

关于张码与 ASCII 码的辨认问题，在本发明软件系统中是不存在的。凡是用到处理中文的函数，所处理的字符一定是张码，决不会是 ASCII 码。同样的，凡是用到其他的函数，所处理的字符一定是 ASCII 码，决不是张码。

本发明采用汉字“笔形检字编码法”，避免数笔划检字的缺点，是对传统的部首检字法作出的重大改进。而且其规则简单易学，使用时无须特殊的训练；利用屏幕显示逐次缩小的检字范围，代替了大键盘，形成改良的中文打字机，使用户能以熟悉的方法用计算机处理中文，只用键盘上 1、2、3、4、5 五个数字键分别代表横、直、斜、点、折五种笔划，除此之外，不用其它符号键以利操作；笔形检字编码法用心语言写成应用程序，能适应日新月异的各种新型计算机，并且也易于把用语言写成的计算机软件改为中文版，节省开发软件的人力物力；以张码作为汉字内存码，便于把汉字资料按笔划多少顺序排列。本发明尤为适于企业管理人员的打印机、文员和领导机关中使用的中文对话式的应用。

说 明 书 附 图

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.